Recommendation for Dealing with "Failed" Control Tests

August 4, 2006
Triggers FG Recommendation #8

PROBLEM STATEMENT: The Irrigated Lands Program (ILP) is responsible for characterizing both irrigation and stormwater runoff from agricultural lands. The ILP has a Technical Issues Committee (TIC) that is working on various issues related to performing watershed ambient water and sediment monitoring. The agricultural community has formed Coalitions for their watersheds so that they have a logical geographical boundary for monitoring, as well as a group of growers to pay for the monitoring. The watersheds range from relatively small (e.g., a water district with 3-5 samples per sampling event) to very large (e.g., Sacramento Valley with ~20 sites per sampling event); most Coalitions are monitoring anywhere from 8 to 12 events per year. Toxicity testing for these Coalitions includes acute *Ceriodaphnia*, acute fathead minnow, and the chronic *Selenastrum* tests.

One of the current topics that the TIC is addressing is how to handle a test (or tests) for which the accompanying Control treatment survival response, growth response, and/or inter-replicate variability does not meet the acceptability criteria established for NPDES testing in the EPA test guidance manuals.

Test	EPA Manual Test Acceptability Criteria
Acute *Ceriodaphnia*	≥90% survival in the Control treatment
Acute Fathead Minnow	≥90% survival in the Control treatment
Chronic *Selenastrum*	mean cell density of 200,000 cells/mL (when tested w/o EDTA) and CV <20% at the Control treatment.

For NPDES testing, the manual indicates that if the test acceptability criteria are not met, the test must be repeated with a newly collected sample.

The initial statement made in both the acute and chronic manual (Section 1.1) indicates that "This manual describes acute (or chronic) toxicity tests for use in the National Pollutant Discharge Elimination System (NPDES) Permits Program". The test conditions and test acceptability criteria within these manuals were established with this in mind. However, the logistical and regulatory framework inherent to the ILP monitoring is very different from most (if not all) NPDES testing situations.

Difference #1: Within the NPDES regulatory framework, a single exceedance of a water quality objective may have significant consequences in terms of fines and/or lawsuits. The need to be prepared for a potential litigation scenario requires a more rigorous QA program that includes re-sampling and/or re-testing, than does a more wide-scale monitoring/characterization program.

Difference #2: For almost all NPDES tests, the consequence of having to re-collect a sample and perform a new test is relatively negligible.

1. In virtually all NPDES cases, there will only have to be one sample (i.e., one "point source" discharge) re-tested, as most NPDES testing is done on a '1 sample at-a-time' basis. If the Control test fails to meet test acceptability criteria, then only one sample will need to be re-tested, at minimal cost for the single test.

- Conversely, the need to test as many as 20+ ILP water samples at-a-time requires that a Control test will be used by multiple samples … if that Control does not meet test acceptability criteria, then ALL of the accompanying ILP samples will need to be re-tested, at significant cost.

2. The sample is easily obtained (grab samples of most point-source discharges can generally be collected within 15 minutes), at very low (if any) extra cost;
   - Conversely, the collection of a new set of ILP water samples can take a two-person team as much as several days, not including the time necessary to prepare sampling plans, mobilization, and de-mobilization.

The potential cost, in both time (i.e., scheduling of events to monitor certain processes (e.g., stormwater runoff) and money to re-sample and/or re-test a set of ILP water samples is extremely high. Given this, the fundamental question before the TIC is: how should we deal with tests whose Control treatment(s) do not meet the NPDES test acceptability criteria?

In answering this question, a new set of important "follow-up" questions and/or issues becomes apparent. In order to answer the more fundamental questions, it will be important to consider these other questions:

- Arguably, the most important consideration in determining how to deal with a "failed" Control test is the question of whether or not the "failed" Control test truly precludes a definitive determination as to the presence or absence of toxicity. The answer to that is most often very straightforward: NO. While less-than-optimal survival or growth responses, or higher-than-optimal variability levels, in the Control treatment are clearly not desirable, they do not preclude an assessment of whether or not a given water sample is toxic. The simple fact-of-the matter is that the vast majority of ILP monitoring water samples do not exhibit any impairment (e.g., reduced survival). It is important to remember that if a sample is toxic, there must be some degree of impairment; if there is no impairment, then ***by definition***, there is no toxicity.

  - For the **acute survival tests**, it is almost always the case that there is 90 -100% survival in the ambient water samples. In these cases, a Control treatment is really not even necessary: the absence of any impairment in the water sample is in-and-of-itself (and by definition) a definitive indication that there is no toxicity.

  - For the **chronic algal growth test**, the vast majority of water samples exhibit algal growth that is markedly greater than the Control treatment response, with the magnitude of this difference typically being greater than 2-fold, and often approaching or exceeding an order of magnitude. Again, it is important to remember that if a sample is toxic, there must be some degree of impairment; if there is no impairment, then ***by definition***, there is no toxicity.

Finally, it is important to consider the situation that existed when the EPA first implemented Whole Effluent Toxicity (WET) testing within the NPDES framework, and first provided the EPA manuals for these toxicity tests (from which the Control test acceptability criteria has been handed down from edition-to edition since). While current Regional Board staff includes

scientists with considerable expertise in the science of aquatic toxicity testing, at the time of the implementation of WET testing and the EPA manuals, the Regional Board staff consisted of primarily engineers and hydrologists. In order to facilitate the use of toxicity tests for regulatory decision-making by individuals lacking the related scientific background, WET is based in large part on a YES/NO type of flow-chart/decision-tree framework.

The situation today is very different, particularly for a program like the ILP which has numerous scientists of national caliber helping to evaluate and interpret the monitoring and toxicity testing data. This wealth of scientific expertise allows us to make decisions and evaluate results on the basis of Good Science and Best Professional Judgement. And in fact, the Phase I study report states: “Best professional judgement will be used in interpretation of results obtained when deviations in the test conditions occurred.”

In fact, a cursory review of the Regional Board’s own Phase 1 Ag Waiver study reveals several deviations from the EPA manual QA measures and test conditions that were apparently made on the basis that ‘the Science’ and “Best Professional Judgement” out-weighed arbitrarily-imposed regulatory test conditions. For example, in the Phase I tests, water samples were aerated prior to testing, whereas the EPA manual states that there is to be no aeration, unless the D.O. concentration falls below certain thresholds. While increasing the ability of the testing lab to achieve levels of survival that meet the test acceptability criteria, aeration also has the potential to strip volatile compounds out of solution and/or to facilitate oxidative degradation of organic compounds. It is important to note that this type of deviation from the EPA manual has a far greater potential to affect the ability of a given toxicity test to accurately assess the presence (or absence) of toxicity (as described in the EPA manual), than does a Control test with 89% survival rather than 90% survival, or an algal test with 25% CV rather than 20% CV.

The EPA manuals are provided as guidance documents for regulators, who are often not aquatic toxicologists, with a "YES/NO" framework for regulating. We are not arguing that application of the EPA manual for ambient water monitoring is inappropriate, but are cautioning all involved to understand that the extension of the manual to other monitoring (e.g., ambient monitoring) should include a thorough understanding of how NPDES permit monitoring and ambient monitoring differ, and therefore should also include the consideration of alternative approaches for dealing with the complex issues involved with ambient monitoring projects based upon “Good Science”.

RECOMMENDATION: In taking into account the above questions, issues, and examples, the resolution of “how to deal with a test (or tests) for which the accompanying Control treatment survival response, growth response, and/or inter-replicate variability does not meet the NPDES test acceptability criteria” seems best addressed via a flow-chart type decision-making structure.

The paradigm that **if a sample is toxic, there must be some degree of impairment; if there is no impairment, then *by definition*, there is no toxicity** should be a fundamental element of this decision-making.

Decision Step 1: If the Control treatment meets all test acceptability criteria, then proceed to statistical analyses for determination of the presence of statically significant reductions in organism survival or algal growth.

Proposed Decision Step 2a: If an acute test of a water sample exhibits 90-100% survival, then by definition, the sample is not toxic, regardless of the Control treatment performance. No further testing is required; test result should be “flagged” to denote <90% survival in the Control treatment.

Proposed Decision Step 2b: If an algal test of a water sample exhibits an algal cell density that is markedly greater than the algal cell density at the Control treatment, then by definition, the sample is not toxic. If the Control test does not meet the NPDES test acceptability criteria, it is proposed that instead of the one-tailed statistical tests (which ask only if the test response for a sample is “less” than the Control), a 2-tailed statistical test be performed. If the results of that test indicate that the algal growth in the water sample is significantly greater than the Control treatment, then the sample should be determined to be not toxic; test result should be “flagged” to denote <200,000 cells/ml of CV>20% survival in the Control treatment.

Proposed Decision Step 3: If the Control treatment does not meet NPDES test acceptability criteria, and the conditions of Steps 2a and 2b are not met, then the water sample in question should be immediately re-tested.

Examples:

- For the **acute survival tests**, it is almost always the case that there is 90 -100% survival in the ambient water samples. In these cases, a Control treatment is really not even necessary: the absence of any impairment in the water sample is in and of itself (and by definition) a definitive indication that there is no toxicity.

Example 1:

Test Treatment	% Survival	Is Sample Toxic?
Control	90	(a passing Control test)
Water Sample	90	By definition, *no*.

Example 2:

Test Treatment	% Survival	Is Sample Toxic?
Control	85	(a "failed" Control test)
Water Sample	90	By definition, *no*.

Example 3:

Test Treatment	% Survival	Is Sample Toxic?
Control	50	(a failed Control test)
Water Sample	90	The answer is still: *no*.

- For the **chronic algal growth test**, the vast majority of water samples exhibit algal growth that is markedly greater than the Control treatment response, with the magnitude of this difference typically being greater than 2-fold, and often approaching or exceeding an order of magnitude. Again, it is important to remember that if a sample is toxic, there must be some degree of impairment; if there is no impairment, then *by definition*, there is no toxicity.

Example 4:

Test Treatment	Cell Density	Variability	Is Sample Toxic?
Control	200,000	20%	(a passing Control test)
Water Sample	200,000		By definition, *no*.

Example 5:

Test Treatment	Cell Density	Variability	Is Sample Toxic?
Control	**190,000**	20%	(a "failed" Control test)
Water Sample	200,000		By definition, *no*.

Example 6:

Test Treatment	Cell Density	Variability	Is Sample Toxic?
Control	200,000	**30%**	(a "failed" Control test)
Water Sample	200,000		The answer is still: *no*.